

Additional File 4. Summary of the effect of mean batch-centering on data generated from published studies.

	SAM Common			
	Richardson <i>et al.</i> 18 basal v Farmer <i>et al.</i> 27 luminal (top 50)		Farmer <i>et al.</i> 16 basal v Richardson <i>et al.</i> 20 non basal-like (top 50)	
	UC	MC	UC	MC
Farmer <i>et al.</i> 16 basal v 27 luminal (top 50)	1	8	2	31
Richardson <i>et al.</i> 18 basal v 20 luminal (top 50)	2	6	4	26
Common to both Farmer <i>et al.</i> and Richardson <i>et al.</i> Luminal v Basal (total 13 probesets)	0	4	0	8

Lists of the top 50 differentially expressed probesets between basal and non basal-like/ luminal tumours were identified within and across datasets, before and after mean batch-centering. SAM Common: for each column two different pairwise comparisons using SAM were performed, and the top 50 probesets identified for each comparison. The number reported is the intersection between two lists. UC = uncorrected. MC = Mean centering correction.